SDS 383D Chapter 1: Preliminaries

Xizewen Han

February 3, 2019

Bayesian Inference in Simple Conjugate Families

(A)

Sol: For $i = 1, \ldots, N$ such that $x_i \sim \text{Bernoulli}(w)$, we have

$$p(x_i|w) = w^{x_i}(1-w)^{1-x_i}$$

Therefore with prior $w \sim \text{Beta}(a, b)$, we have the posterior distribution

$$\begin{aligned}
p(w|x_{1:N}) &\propto p(x_{1:N}|w) \cdot p(w) \\
&= \prod_{i=1}^{N} p(x_i|w) \cdot p(w) \\
&= w^{\sum x_i}(1-w)^{N-\sum x_i} \cdot \frac{1}{B(a,b)} w^{a-1}(1-w)^{b-1} \\
&\propto w^{\sum x_i + a - 1}(1-w)^{N - \sum x_i + b - 1}
\end{aligned}$$

Note that the last expression is the kernel of $\text{Beta}(\sum_{i=1}^{N} x_i + a, N - \sum_{i=1}^{N} x_i + b)$ distribution, which by adding the normalizing term we'd have this as the posterior distribution. ■

(B)

Sol: For such transformation, which is one-to-one and onto, we have the inverse as

$$\begin{cases} x_1 = y_1 y_2, \\ x_2 = y_2 - y_1 y_2 \end{cases}$$

Therefore, we have the Jacobian as

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{bmatrix} = y_2 \geq 0$$

thus the joint distribution is

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}\big(x_1(y_1,y_2), x_2(y_1,y_2)\big)|J| \\
&= f_{X_1}\big(x_1(y_1,y_2)\big) f_{X_2}\big(x_2(y_1,y_2)\big)|J| \\
&= \frac{1}{\Gamma(a_1)}(y_1y_2)^{a_1-1}e^{-y_1y_2} \cdot \frac{1}{\Gamma(a_2)}(y_2 - y_1y_2)^{a_2-1}e^{-y_2+y_1y_2} \cdot y_2 \\
&= \underbrace{\frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1}}_{g(y_1),\text{ kernel of Beta}(a_1,a_2)} \underbrace{y_2^{a_1+a_2-1}e^{-y_2}}_{h(y_2),\text{ kernel of Gamma}(a_1+a_2,1)}
\end{aligned}
$$

Since the joint distribution of Y_1 and Y_2 can be factored into two functions, $g(y_1)$ and $h(y_2)$, by Lemma 4.2.7 of Casella and Berger's Statistical Inference, as well as the form of each function noted above in the last expression, we conclude that Y_1 and Y_2 are independent random variables such that

$$Y_1 \sim \text{Beta}(a_1, a_2)$$

and

$$Y_2 \sim \text{Gamma}(a_1 + a_2, 1)$$

Based on such conclusion, we propose the following method to simulate a Beta(a_1, a_2) random variable:

- 1st step: simulate from two Gamma random variables: $X_1 \sim \text{Gamma}(a_1, 1)$ and $X_2 \sim \text{Gamma}(a_2, 1)$
- 2nd step: calculate $Y_1 = \frac{X_1}{X_1+X_2}$

Such $Y_1 \sim \text{Beta}(a_1, a_2)$. ■

(C)

Sol: With $X_i \sim \mathcal{N}(\theta, \sigma^2)$ for $i = 1, \ldots, N$, in which θ is unknown and σ^2 is known, and $\theta \sim \mathcal{N}(m, v)$, we have the posterior distribution

$$
\begin{aligned}
p(\theta|x_{1:N}) &\propto p(\theta) \cdot p(x_{1:N}|\theta) \\
&= \frac{1}{\sqrt{2\pi v}} e^{-\frac{(\theta-m)^2}{2v}} \cdot \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}} \\
&\propto \exp\Big(-\frac{(nv+\sigma^2)\theta^2 - 2\big(m\sigma^2 + v(\sum x_i)\big)}{2\sigma^2 v}\Big)
\end{aligned}
$$

By noting the kernel, we recognize the posterior distribution as

$$\mathcal{N}\Big(\frac{m\sigma^2 + v(\sum_{i=1}^N x_i)}{\sigma^2 + nv}, \frac{\sigma^2 v}{\sigma^2 + nv}\Big)$$

The variance term can be re-written as

$$\frac{\sigma^2 v}{\sigma^2 + nv} = \frac{1}{\frac{1}{v} + \frac{n}{\sigma^2}}$$

Note the denominator is the precision, and it's the sum of data precision and prior prevision − a good characteristic of posterior of normal-normal model. ■

(D)

Sol: With $X_i \sim \mathcal{N}(\theta, 1/w)$ for $i = 1, \dots, N$, in which θ is known and precision w is unknown, and $w \sim \text{Gamma}(a, b)$, we have the posterior distribution

$$\begin{aligned} p(w|x_{1:N}) &\propto p(w) \cdot \prod_{i=1}^{N} p(x_i|\theta, w) \\ &\propto w^{a-1} e^{-\frac{1}{b}w} w^{\frac{N}{2}} e^{-\frac{\sum(x_i-\theta)^2}{2}w} \\ &= w^{\frac{N}{2}+a-1} e^{-\left(\frac{1}{b}+\frac{\sum(x_i-\theta)^2}{2}\right)w} \end{aligned}$$

By recognizing the kernel, we have the posterior distribution

$$w \sim \text{Gamma}\Big(\frac{N}{2} + a, \frac{1}{b} + \frac{\sum(x_i - \theta)^2}{2}\Big)$$

Since w is a re-parameterization of σ^2, we use change of variables directly on posterior pdf to derive the posterior of σ^2. We have the transformation $g(w) = \frac{1}{\sigma^2}$, which is one-to-one and onto, and

$$\frac{d}{d\sigma^2} g^{-1}(\sigma^2) = \frac{d}{d\sigma^2} (\sigma^2)^{-1} = -(\sigma^2)^{-2}$$

Therefore, for a general case $w \sim \text{Gamma}(a, b)$, we have

$$p(\sigma^2) = p\big(g^{-1}(\sigma^2)\big) \left| \frac{d}{d\sigma^2} g^{-1}(\sigma^2) \right| = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a-1)} e^{-\frac{b}{\sigma^2}} (\sigma^2)^{-2} = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}$$

i.e. $\sigma^2 \sim IG(a, b)$. Thus for our specific case, the posterior distribution of σ^2 is

$$\sigma^2 \sim IG\Big(\frac{N}{2} + a, \frac{1}{b} + \frac{\sum(x_i - \theta)^2}{2}\Big)$$

■

(E)

Sol: With known idiosyncratic variances for normal likelihood, the posterior density for unknown common mean θ is

$$\begin{aligned} p(\theta|x_{1:N}) &\propto \mathbb{P}(\theta) \cdot \prod_{i=1}^{N} p(x_i|\theta) \\ &= \frac{1}{\sqrt{2\pi v}} e^{-\frac{(\theta-m)^2}{2v}} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\sum \frac{(x_i-\theta)^2}{\sigma_i{}^2}} \\ &\propto e^{-\frac{1}{2}\left(\left(\frac{1}{v}+\sum\frac{1}{\sigma_i{}^2}\right)\theta^2 - 2\left(\frac{m}{v}+\sum\frac{x_i}{\sigma_i{}^2}\right)\theta\right)} \end{aligned}$$

We can thus directly write out the posterior distribution for θ as

$$\mathcal{N}\Big(\frac{\frac{1}{v}m + \sum_{i=1}^{N} \frac{1}{\sigma_i{}^2} x_i}{\frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i{}^2}}, \frac{1}{v} + \sum_{i=1}^{N} \frac{1}{\sigma_i{}^2}\Big)$$

notice that the mean term is expressed in the form as a weighted average of the observations x_i's and the prior mean m.

A later note: the weighted average part is more obvious if we work with precision instead of variance: denote prior precision as $w = \frac{1}{v}$ and idiosyncratic precision for each observation as $w_i = \frac{1}{{\sigma_i}^2}$, we have the posterior distribution in the form of

$$\mathcal{N}\Big(\frac{w \cdot m + \sum_{i=1}^N w_i \cdot x_i}{w + \sum_{i=1}^N w_i}, w + \sum_{i=1}^N w_i\Big)$$

■

(F)

Sol: Using the parameterization from the question sheet, we compute the marginal distribution of x as

$$\begin{aligned}
p(x) &= \int_w p(x|w)p(w)dw \\
&= \int_w (\frac{w}{2\pi})^{1/2} e^{-\frac{w}{2}(x-m)^2} \frac{(b/2)^{a/2}}{\Gamma(a/2)} w^{a/2-1} e^{-\frac{b}{2}w} \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \cdot \int_0^\infty \underbrace{w^{\frac{a+1}{2}-1} e^{-\left(\frac{b}{2}+\frac{(x-m)^2}{2}\right)w}}_{\text{kernel of Gamma}(\frac{a+1}{2}, \frac{b+(x-m)^2}{2})} \quad dw \\
&= \frac{(b/2)^{a/2}}{\sqrt{2\pi}\Gamma(a/2)} \frac{\Gamma(\frac{a+1}{2})}{\left(\frac{b+(x-m)^2}{2}\right)^{\frac{a+1}{2}}} \cdot 1 \\
&= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})} \frac{1}{\sqrt{\pi b}} \Big(1 + \frac{(x-m)^2}{b}\Big)^{-\frac{a+1}{2}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})} \frac{1}{\sqrt{a\pi}(b/a)^{1/2}} \Big(\frac{a + \left(\frac{x-m}{(b/a)^{1/2}}\right)^2}{a}\Big)^{-\frac{a+1}{2}}
\end{aligned}$$

Note that the last expression is exactly the pdf of t location-scale distribution with $d = a$ degrees of freedom, center m, and scale parameter $(b/a)^{1/2}$.

Later note: this question demonstrates a useful conclusion – if we want to sample from t-distribution, which has heavy tails, we don't have to sample directly; instead, we first sample precision from Gamma distribution, then sample observation from normal distribution given that precision. The resulting sample is a random variable with t-distribution. ■

The Multivariate Normal Distribution

(A)

Sol: For the two sub-problems, we do them by using the definition given by the question and directly applying matrix operation rules:

(1)

$$\begin{aligned}
cov(\boldsymbol{x}) &= \mathbb{E}\big((\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\big) \\
&= \mathbb{E}\big((\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}^T-\boldsymbol{\mu}^T)\big) \\
&= \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T - \boldsymbol{x}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{x}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T) \\
&= \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T) - \mathbb{E}(\boldsymbol{x})\boldsymbol{\mu}^T - \boldsymbol{\mu}\mathbb{E}(\boldsymbol{x}^T) + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \text{ (using the fact that } \mathbb{E}(\boldsymbol{x}) = \mu) \\
&= \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}$$

The last expression is what's asked to show. ■

(2)

Sol: The mean of $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ is:

$$\mathbb{E}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) = \boldsymbol{A}\mathbb{E}(\boldsymbol{x}) + \mathbb{E}(\boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}$$

Therefore, we have

$$\begin{aligned}
cov(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) &= \mathbb{E}\Big(\big((\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) - (\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})\big)\big((\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) - (\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})\big)^T\Big) \\
&= \mathbb{E}\Big(\big(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})\big)\big(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})\big)^T\Big) \\
&= \mathbb{E}\big(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{A}^T\big) \\
&= \boldsymbol{A}\mathbb{E}\big((\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\big)\boldsymbol{A}^T \\
&= \boldsymbol{A}cov(\boldsymbol{x})\boldsymbol{A}^T \text{ (the middle term is derived by the definition of covariance given by the question)}
\end{aligned}$$

The last expression is what's asked to show. ■

(B)

Sol: Since for all z_i's such that $i = 1, \ldots, p$, they are independent, the pdf of $\boldsymbol{z} = (z_1, \ldots, z_p)^T$, being the joint distribution of z_i's, is the multiplication of their individual densities:

$$\begin{aligned}
f_{\boldsymbol{Z}}(\boldsymbol{z}) &= \prod_{i=1}^{p} f(z_i) \\
&= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}}\exp(-\frac{{z_i}^2}{2}) \\
&= \frac{1}{(2\pi)^{p/2}} \cdot \exp\big(-\frac{1}{2}\sum {z_i}^2\big) \\
&= \frac{1}{(2\pi)^{p/2}} \cdot \exp\big(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\big)
\end{aligned}$$

The mgf of $\boldsymbol{z}$ is given by

$$
\begin{aligned}
\mathbb{E}(e^{\boldsymbol{t^T z}}) &= \int_{z_1} \cdots \int_{z_p} e^{\boldsymbol{t^T z}} f_{\boldsymbol{Z}}(\boldsymbol{z}) dz_1 \dots dz_p \\
&= \int_{z_1} \cdots \int_{z_p} e^{\sum t_i z_i} f_{\boldsymbol{Z}}(\boldsymbol{z}) dz_1 \dots dz_p \\
&= \int_{z_1} \cdots \int_{z_p} \prod_{i=1}^{p} e^{t_i z_i} \prod f_{Z_i}(z_i) dz_1 \dots dz_p \\
&= \prod_{i=1}^{p} \int_{z_i} e^{t_i z_i} f_{Z_i}(z_i) dz_i \text{ (using the fact that } z_i\text{'s are independent)} \\
&= \prod_{i=1}^{p} \exp(m t_i + v {t_i}^2/2) \text{ (given by question)} \\
&= \prod_{i=1}^{p} \exp(0 \cdot t_i + 1 \cdot {t_i}^2/2) \text{ (given by question)} \\
&= \prod_{i=1}^{p} \exp({t_i}^2/2) \\
&= \exp(\frac{{t_i}^2}{2}) \\
&= \exp(\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t})
\end{aligned}
$$

$\blacksquare$

(C)

Sol: To show that if $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, its mgf is of the given form, we use the definition given in this problem: for a multivariate normal variable $\boldsymbol{x}$, $z = \boldsymbol{a^T x}$ is univariate normal.

It's expectation is

$$\mathbb{E}(z) = \mathbb{E}(\boldsymbol{a^T x}) = \boldsymbol{a^T}\mathbb{E}(\boldsymbol{x}) = \boldsymbol{a^T \mu}$$

and

$$var(z) = cov(\boldsymbol{a^T x}) = \boldsymbol{a^T} cov(\boldsymbol{a^T x})\boldsymbol{a} = \boldsymbol{a^T \Sigma a}$$

We know the mgf of a univariate normal variable as

$$M_z(t) = (\boldsymbol{a^T \mu})t + (\boldsymbol{a^T \Sigma a})\frac{t^2}{2} = (t\boldsymbol{a})^T\boldsymbol{\mu} + \frac{(t\boldsymbol{a})^T\boldsymbol{\Sigma}(t\boldsymbol{a})}{2}$$

meanwhile

$$M_z(t) = \mathbb{E}(e^{zt}) = \mathbb{E}(e^{(\boldsymbol{a^T x})t}) = \mathbb{E}(e^{(t\boldsymbol{a})^T\boldsymbol{x}}) = M_{\boldsymbol{x}}(t\boldsymbol{a})$$

therefore, denoting $\boldsymbol{t} = t\boldsymbol{a}$, we have the mgf of multivariate normal variable $\boldsymbol{x}$ as

$$M_{\boldsymbol{x}}(\boldsymbol{t}) = \boldsymbol{t^T \mu} + \frac{\boldsymbol{t^T \Sigma t}}{2}$$

For the other direction, i.e. with the given form of pdf, such variable is multivariate normal, we again use the definition given in this particular question s.t. every linear combination of a multivariate normal random variable is univariate normal:

for all vectors $\boldsymbol{a}$ such that $z = \boldsymbol{a}^T\boldsymbol{x}$, we have its moment generate function as

$$\begin{aligned}
M_z(t) &= M_{\boldsymbol{a}^T\boldsymbol{x}}(t) \\
&= \mathbb{E}\big(e^{t\cdot(\boldsymbol{a}^T\boldsymbol{x})}\big) \\
&= \mathbb{E}\big(e^{(t\boldsymbol{a})^T\cdot\boldsymbol{x}}\big) \\
&= M_{\boldsymbol{x}}(t\boldsymbol{a}) \\
&= \exp\big((t\boldsymbol{a})^T\boldsymbol{\mu} + \frac{1}{2}(t\boldsymbol{a})^T\boldsymbol{\Sigma}(t\boldsymbol{a})\big) \\
&= \exp\big(t(\boldsymbol{a}^T\boldsymbol{\mu}) + \frac{1}{2}t^2(\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a})\big) \\
&= \exp\big(m\cdot t + v\cdot t^2/2\big)
\end{aligned}$$

the last expression is exactly the form of the mgf of univariate normal distribution $\mathcal{N}(m, v)$, in which $m = \boldsymbol{a}^T\boldsymbol{\mu}$, $v = \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}$, i.e. z has a $\mathcal{N}(m, v)$ distribution. Therefore, according to the definition given by the question, $\boldsymbol{x}$ is multivariate normal. ■

(D)

Sol: The question defines

$$\boldsymbol{x} = \boldsymbol{L}\boldsymbol{z} + \boldsymbol{\mu}$$

in which

$$\boldsymbol{z} = (z_1, \ldots, z_p)^T$$

s.t. each z_i for $i = 1, \ldots, p$ are i.i.d. $\mathcal{N}(0, 1)$ distribution.

I want to solve this problem with two methods, with the intention of going through different parts of mathematical statistics knowledge as a review:

Method 1

Since z_i's are mutually independent, by Corollary 4.6.10 from Casella and Berger's *Statistical Inference* each of their linear combinations is univariate normal, thus $\boldsymbol{z}$ is a multivariate normal random variable. Its pdf can be derived by

$$p(\boldsymbol{z}) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-{z_i}^2/2} = \frac{1}{(2\pi)^{p/2}} e^{-\boldsymbol{z}^T\boldsymbol{z}/2}$$

which is the form of a standard multivariate normal variable's pdf – we just proved the claim/definition given from (B). We directly write out our result from (B): the mgf of $\boldsymbol{z}$ is

$$M_z(\boldsymbol{t}) = \exp\big(\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}\big)$$

Now we show $\boldsymbol{x}$ is a multivariate normal variable: using change of variable, we have

$$\boldsymbol{x} = g(\boldsymbol{z}) = \boldsymbol{L}\boldsymbol{z} + \boldsymbol{\mu}$$

since $\boldsymbol{L}$ is full rank, it's invertible, thus we have

$$\boldsymbol{z} = g^{-1}(\boldsymbol{x}) = \boldsymbol{L}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$$

Therefore we have the Jacobian

$$J = \Big|\frac{\partial \boldsymbol{r}(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big| = \Big|\frac{\partial \boldsymbol{g^{-1}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big| = \big|\boldsymbol{L}^{-1}\big| = \big|\boldsymbol{L}\big|^{-1}$$

The pdf of $\boldsymbol{x}$ is thus

$$\begin{aligned}
p(\boldsymbol{x}) &= \frac{1}{(2\pi)^{p/2}} e^{-\boldsymbol{(L^{-1}(x-\mu))^T(L^{-1}(x-\mu))/2}} \cdot |J| \\
&= (2\pi)^{-p/2}\big|\boldsymbol{L}\big|^{-1}\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{L}^{-1})^T(\boldsymbol{L}^{-1})(\boldsymbol{x}-\boldsymbol{\mu})\big) \\
&= (2\pi)^{-p/2}\big|\boldsymbol{L}\big|^{-1}\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{L}^T)^{-1}(\boldsymbol{L}^{-1})(\boldsymbol{x}-\boldsymbol{\mu})\big) \\
&= (2\pi)^{-p/2}\big|\boldsymbol{L}\big|^{-1}\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{L}\boldsymbol{L}^T)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)
\end{aligned}$$

Since $\boldsymbol{L}$ is full rank, $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{L}\boldsymbol{I}_{p\times p}\boldsymbol{L}^T$ is positive definite, in which $\boldsymbol{I}_{p\times p}$ is identity matrix (thus positive definite). Note that $p(\boldsymbol{x})$ is exactly the form of the pdf of $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{L}\boldsymbol{L}^T)$ distribution.

Method 2

I also want to use (a simple manipulation on) mgf to solve this question:

Since we've already derived the mgf of $\boldsymbol{z}$ as a standard multivariate normal random variable, i.e.

$$M_z(\boldsymbol{t}) = \exp\big(\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}\big)$$

we can directly write out the mgf of $\boldsymbol{x}$ as

$$\begin{aligned}
M_{\boldsymbol{x}}(\boldsymbol{t}) &= \mathbb{E}(e^{t^T x}) \\
&= \mathbb{E}(e^{t^T(\boldsymbol{Lz}+\boldsymbol{\mu})}) \\
&= \exp(\boldsymbol{t^T\mu})\mathbb{E}\big((\boldsymbol{L^T t})^T\boldsymbol{z}\big) \\
&= \exp(\boldsymbol{t^T\mu})M_z(\boldsymbol{L^T t}) \\
&= \exp(\boldsymbol{t^T\mu})\exp\big(\frac{1}{2}(\boldsymbol{L^T t})^T(\boldsymbol{L^T t})\big) \\
&= \exp\big(\boldsymbol{t^T\mu} + \frac{1}{2}\boldsymbol{t^T L L^T t}\big)
\end{aligned}$$

thus by our conclusion from (C), a random variable with this form of mgf has a $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{LL^T})$ distribution.

To compute the expected value and covariance matrix of $\boldsymbol{x}$, we take the first and second derivative of mgf at $t = 0$, respectively:

$$\begin{aligned}
\mathbb{E}(\boldsymbol{x}) &= \frac{\partial M_{\boldsymbol{x}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\Big|_{t=0} \\
&= \frac{\partial}{\partial \boldsymbol{t}} \int_{\boldsymbol{x}} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \frac{\partial}{\partial \boldsymbol{t}} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \frac{\partial}{\partial \boldsymbol{t}} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big)(\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \exp\big(\boldsymbol{0}\big)(\boldsymbol{\mu}) f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \boldsymbol{\mu} \int_{\boldsymbol{x}} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \boldsymbol{\mu}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(\boldsymbol{x^T}\boldsymbol{x}) &= \frac{\partial^2 M_{\boldsymbol{x}}(\boldsymbol{t})}{\partial \boldsymbol{t}^T \boldsymbol{t}}\Big|_{t=0} \\
&= \frac{\partial}{\partial \boldsymbol{t}^T} \frac{\partial M_{\boldsymbol{x}}(\boldsymbol{t})}{\partial \boldsymbol{t}}\Big|_{t=0} \\
&= \frac{\partial}{\partial \boldsymbol{t}^T} \int_{\boldsymbol{x}} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big)(\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \frac{\partial}{\partial \boldsymbol{t}^T} \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big)(\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \Big(\exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big)(\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t})^T(\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}) + \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t^T}\boldsymbol{L}\boldsymbol{L^T}\boldsymbol{t}\big)(\boldsymbol{L}\boldsymbol{L^T})\Big) f(\boldsymbol{x}) d\boldsymbol{x}\Big|_{t=0} \\
&= \int_{\boldsymbol{x}} \Big(\exp\big(\boldsymbol{0}\big)(\boldsymbol{\mu})^T(\boldsymbol{\mu}) + \exp\big(\boldsymbol{0}\big)(\boldsymbol{L}\boldsymbol{L^T})\Big) f(\boldsymbol{x}) d\boldsymbol{x} \\
&= (\boldsymbol{\mu^T}\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}) \int_{\boldsymbol{x}} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \boldsymbol{\mu^T}\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T}
\end{aligned}$$

therefore

$$cov(\boldsymbol{x}) = \mathbb{E}(\boldsymbol{x^T}\boldsymbol{x}) - \mathbb{E}(\boldsymbol{x})^T\mathbb{E}(\boldsymbol{x}) = \boldsymbol{\mu^T}\boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L^T} - \boldsymbol{\mu}^T\boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{L^T}$$

■

(E)

Sol: For the "only if" direction, i.e. for multivariate normal variable $\boldsymbol{x}$, we could argue in the following manner: according to Casella and Berger's arguments under Theorem 2.3.12 in *Statistical Inference* about the uniqueness of Laplace transforms, each mgf can only have one pdf, i.e. one distinct distribution being mapped to it. Therefore, the logic to solve this problem is that if we can find an affine transformation of independent univariate normal variables such that they have the same mgf as our desired multivariate normal variable's mgf, we can conclude that the transformation and the multivariate variable are from the same distribution, i.e. the multivariate normal variable can be expressed as an affine transformation of a standard multivariate normal variable.

We had from Question (C) such that for a $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random variable $\boldsymbol{x}$, it has mgf

$$M_{\boldsymbol{x}}(\boldsymbol{t}) = \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{\boldsymbol{t^T\Sigma t}}{2}\big)$$

Meanwhile, we've found in (D) s.t. an affine transformation $\boldsymbol{Lz}+\boldsymbol{\mu}$ to standard multivariate normal variable $\boldsymbol{z}$ has mgf

$$M_{\boldsymbol{Lz}+\boldsymbol{\mu}}(\boldsymbol{t}) = \exp\big(\boldsymbol{t^T}\boldsymbol{\mu} + \frac{\boldsymbol{t^T LL^T t}}{2}\big)$$

Therefore, by simply observing the forms of these two mgf's we can find, by setting $\boldsymbol{L}$ s.t. $\boldsymbol{LL^T} = \boldsymbol{\Sigma}$, we'd have the same form of mgf. We can achieve this by performing Cholesky decomposition on $\boldsymbol{\Sigma}$: since $\boldsymbol{\Sigma}$ is positive definite, we can find $\boldsymbol{L}$ s.t. $\boldsymbol{\Sigma} = \boldsymbol{LL^T}$.

We've thus found such affine transformation that $\boldsymbol{x} = \boldsymbol{Lz} + \boldsymbol{\mu}$, in which $\boldsymbol{\mu}$ is the mean of $\boldsymbol{x}$ and $\boldsymbol{L}$ is the component of Cholesky decomposition of covariance matrix of $\boldsymbol{x}$.

With this conclusion, I propose the following algorithm to simulate a $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ variable $\boldsymbol{x}$:

(1) simulate $\boldsymbol{z}$ from $\mathcal{MVN}(\boldsymbol{0}, \boldsymbol{I})$ distribution

(2) perform Cholesky decomposition on covariance matrix $\boldsymbol{\Sigma}$ of desired multivariate normal variable:

$$\boldsymbol{\Sigma} = \boldsymbol{LL^T}$$

(3) compute $\boldsymbol{x} = \boldsymbol{Lz} + \boldsymbol{\mu}$ – such $\boldsymbol{x}$ is a sample from $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. ■

(F)

Sol: We first note that a matrix quadratic form $Q(\boldsymbol{x} - \boldsymbol{\mu})$ has the form

$$Q(\boldsymbol{x} - \boldsymbol{\mu}) = (\boldsymbol{x} - \boldsymbol{\mu})^T (A\boldsymbol{x} - \boldsymbol{\mu})$$

in which $\boldsymbol{A}$ is a symmetric matrix.

We've actually solved this problem in our solution for (D) under Method 1 – we applied change of variable on standard multivariate normal variable $\boldsymbol{z}$ to derive the pdf of $\boldsymbol{x} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{LL}^T)$. We briefly go over the steps again here:

The pdf of the standard multivariate normal variable $\boldsymbol{z}$ is

$$p(\boldsymbol{z}) = \frac{1}{(2\pi)^{p/2}} e^{-\boldsymbol{z^T z}/2}$$

From (E) we have that $\boldsymbol{x} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be expressed by $\boldsymbol{z}$ as

$$\boldsymbol{x} = \boldsymbol{Lz} + \boldsymbol{\mu}$$

in which $\boldsymbol{LL^T} = \boldsymbol{\Sigma}$ is the Cholesky decompositon of $\boldsymbol{\Sigma}$.

Applying change of variable:

$$g(\boldsymbol{z}) = \boldsymbol{x} = \boldsymbol{L}\boldsymbol{z} + \boldsymbol{\mu}$$

note that $\boldsymbol{L}$ here is invertible, thus we have

$$\boldsymbol{z} = g^{-1}(\boldsymbol{x}) = \boldsymbol{L}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$$

Therefore we have the Jacobian

$$J = \left|\frac{\partial \boldsymbol{r}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right| = \left|\frac{\partial \boldsymbol{g^{-1}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right| = \left|\boldsymbol{L}^{-1}\right| = \left|\boldsymbol{L}\right|^{-1}$$

The pdf of $\boldsymbol{x}$ is thus

$$\begin{aligned} p(\boldsymbol{x}) &= \frac{1}{(2\pi)^{p/2}} e^{-\boldsymbol{(L^{-1}(x-\mu))^T(L^{-1}(x-\mu))/2}} \cdot |J| \\ &= (2\pi)^{-p/2}\left|\boldsymbol{L}\right|^{-1}\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{L}\boldsymbol{L}^T)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big) \end{aligned}$$

Note that $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{\Sigma}$ which is a symmetric matrix, thus $(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{L}\boldsymbol{L}^T)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is a quadratic form. We thus can express the pdf as

$$p(\boldsymbol{x}) = C \cdot \exp\big(-\frac{1}{2}Q(\boldsymbol{x}-\boldsymbol{\mu})\big)$$

in which

$$C = (2\pi)^{-p/2}\left|\boldsymbol{L}\right|^{-1}$$

and

$$Q(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

$\blacksquare$

(G)

Sol: To solve this problem, we first compute the mgf of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x_1} + \boldsymbol{B}\boldsymbol{x_2}$, then use the "if-and-only-if" relationship between multivariate normal distribution and mgf that we proved in (C) to read out the distribution of $\boldsymbol{y}$ with corresponding parameters.

We have the mgf of $\boldsymbol{y}$ as

$$
\begin{aligned}
M_{\boldsymbol{y}}(\boldsymbol{t}) &= \mathbb{E}(e^{\boldsymbol{t^T y}}) \\
&= \mathbb{E}(e^{\boldsymbol{t^T(Ax_1+Bx_2)}}) \\
&= \mathbb{E}(e^{\boldsymbol{t^T(Ax_1)}}e^{\boldsymbol{t^T(Bx_2)}}) \\
&= \mathbb{E}(e^{\boldsymbol{(A^Tt)^Tx_1}}e^{\boldsymbol{(B^Tt)^Tx_2}}) \\
&= \int_{\boldsymbol{x_1}}\int_{\boldsymbol{x_2}} e^{\boldsymbol{(A^Tt)^Tx_1}}e^{\boldsymbol{(B^Tt)^Tx_2}} f(\boldsymbol{x_1},\boldsymbol{x_2})d\boldsymbol{x_1}d\boldsymbol{x_2} \\
&= \int_{\boldsymbol{x_1}}\int_{\boldsymbol{x_2}} e^{\boldsymbol{(A^Tt)^Tx_1}}e^{\boldsymbol{(B^Tt)^Tx_2}} f(\boldsymbol{x_1})f(\boldsymbol{x_2})d\boldsymbol{x_1}d\boldsymbol{x_2} \text{ } (\boldsymbol{x_1} \text{ and } \boldsymbol{x_2} \text{ are independent}) \\
&= \int_{\boldsymbol{x_1}} e^{\boldsymbol{(A^Tt)^Tx_1}} f(\boldsymbol{x_1})d\boldsymbol{x_1}\int_{\boldsymbol{x_2}} e^{\boldsymbol{(B^Tt)^Tx_2}} f(\boldsymbol{x_2})d\boldsymbol{x_2} \\
&= M_{\boldsymbol{x_1}}(\boldsymbol{A^Tt})\cdot M_{\boldsymbol{x_2}}(\boldsymbol{B^Tt}) \\
&= \exp(\boldsymbol{t^T\mu_1} + \frac{\boldsymbol{t^T\Sigma_1 t}}{2})\cdot\exp(\boldsymbol{t^T\mu_2} + \frac{\boldsymbol{t^T\Sigma_2 t}}{2}) \\
&= \exp\Big(\boldsymbol{t^T}(\boldsymbol{A\mu_1}+\boldsymbol{B\mu_2}) + \frac{\boldsymbol{t^T}(\boldsymbol{A\Sigma_1A^T}+\boldsymbol{B\Sigma_2B^T})\boldsymbol{t}}{2}\Big)
\end{aligned}
$$

This is almost the mgf of a multivariate normal variable, except that we need to justify that $(\boldsymbol{A\Sigma_1A^T} + \boldsymbol{B\Sigma_2B^T})$ is positive definite so that it can be a covariance matrix:

Since both $\boldsymbol{A}$ and $\boldsymbol{B}$ have full column rank, and both $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are positive definite since they are both covariance matrices, $\boldsymbol{A\Sigma_1A^T}$ and $\boldsymbol{B\Sigma_2B^T}$ are each positive definite. The sum of two positive definite matrices (with same dimension) is positive definite too: denote two p.d. matrices as $\boldsymbol{M}$ and $\boldsymbol{N}$, s.t.

$$\boldsymbol{u^TMu} > 0$$

and

$$\boldsymbol{u^TNu} > 0$$

for all nonzero vectors $\boldsymbol{u}$, then

$$\boldsymbol{u^T(M+N)u} = \boldsymbol{u^TMu} + \boldsymbol{u^TNu} > 0$$

Therefore, $(\boldsymbol{A\Sigma_1A^T} + \boldsymbol{B\Sigma_2B^T})$ is positive definite. We thus have

$$\boldsymbol{y} \sim \mathcal{MVN}\Big(\boldsymbol{A\mu_1}+\boldsymbol{B\mu_2}, \boldsymbol{A\Sigma_1A^T}+\boldsymbol{B\Sigma_2B^T}\Big)$$

■

Conditionals and Marginals

(A)

Sol:

■